Seaborn

Author: Siri Duddella Date: May 14, 2025

1 Basic Aim

Seaborn is a Python library that extends matplotlib to improve and provide better tools for statistical data visualization. It provides simpler mechanisms for calculating errors, plotting multiple variables, etc. that may take extra steps in its parent library, matplotlib.

2 Motivation

My motivation for selecting this package is that it's a package that is used very frequently in both the computer science and physics industries for data visualization, which are the two fields I'm currently majoring in.

3 & 4 History and Contributions

Seaborn was initially published in 2012 by Michael Waskom. The library builds on matplotlib and pandas to offer better statistical data visualization tools for processing large data effectively. It's comparable to tools likeggplot2, but instead of focusing on producing arbitrary visualizations, it focuses on developing specific, aesthetic, figures for publication.

This library is still maintained by Waskom on GitHub and is currently on v0.13.2. While there isn't an explicit list of instructions to contribute to the library, the author provides a method to submit bugs a user encounters to an issue tracker on GitHub. For the purpose of this report, I used v0.12.2 (Released December 2022) as shown by this part in my Jupyter Notebook:

```python
import seaborn as sns
print(sns.__version__)
```

5 & 6 Installation and Usage

I installed this package using a simple `%pip install seaborn` in the top cell of my Jupyter Notebook. The package is really easy to install and easy to learn and use since it works so well with pandas DataFrame inputs.

7 Source Code

The source code is still available on GitHub: https://github.com/mwaskom/seaborn.

8 Usage by Other Packages

This package isn't explicitly used by other packages.

9 Usage in Code

This code is used in either python scripts or Jupyter Notebooks.

10 Examples Using the Code

The provided Jupyter Notebook shows one particular example of using seaborn to do a statistical analysis of airline delays.

11 Figures

Seaborn is primarily used to generate figures for data visualization, as it's an extension of matplotlib. It still uses aspects of matplotlib.pyplot to structure its figures.

12 Figure
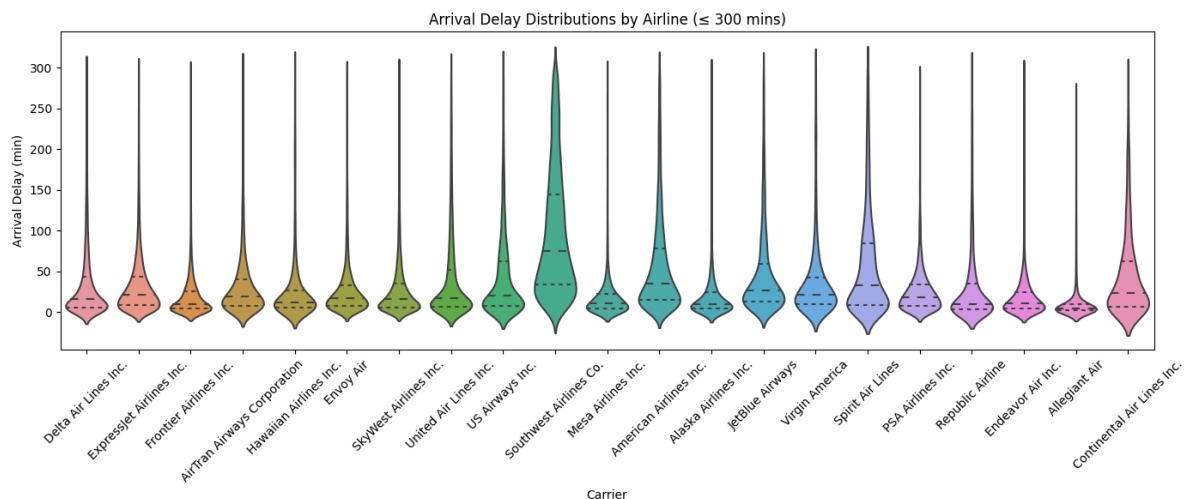
The figure uses `sns.violinplot()` to plot violin plots of each airline and the distribution of its delays using data from January 2011 to December 2020 and clipping extreme delay values (>300 mins). Violin Plots are one of the unique components to seaborn that make it more versatile than matplotlib for its ability to visualize categorical data effectively.

13 Base Language

This package is purely based on Python. It's based on Python libraries as shown by `%pip show seaborn` which prints the dependencies of the library: matplotlib, numpy, and pandas.

14 & 15 Input and Output

The input to this package is usually a Pandas DataFrame but can also be other forms of input that follow either long-form or wide-form data rules.

The output to this package is usually a figure of some sort that visualizes the input dataset in the desired method/plot.

16 & 17 Testing & Confidence

The library offers unit tests which can be run using `make test` in the source directory to generate a coverage report.

Running the unit tests can help with a users confidence in the code's results. Additionally, the produced figure can be compared with expected trends, if available.

18 Dependencies

As stated previously, `%pip show seaborn` lists the dependencies as matplotlib, numpy, and pandas.

19 Documentation

The documentation is provided on the library's website and it was sufficient to understand how to use the library.

20 Preferred Citation Method

In APA:

> Waskom, M. L., (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021, https://doi.org/10.21105/joss.03021.

In BibTex:

> @article{Waskom2021, doi = {10.21105/joss.03021}, url = {https://doi.org/10.21105/joss.03021}, year = {2021}, publisher = {The Open Journal}, volume = {6}, number = {60}, pages = {3021}, author = {Michael L. Waskom}, title = {seaborn: statistical data visualization}, journal = {Journal of Open Source Software} }

21 References

ASCL: https://ascl.net/2012.015 GitHub: https://github.com/mwaskom/seaborn seaborn paper: https://ui.adsabs.harvard.edu/abs/2021JOSS....6.3021W/abstract seaborn website: https://seaborn.pydata.org/index.html

22 Two Other Papers

The ASCL Record lists these 2 papers:

1. https://ui.adsabs.harvard.edu/abs/2019ApJ...883L..42F/abstract
2. https://ui.adsabs.harvard.edu/abs/2021MNRAS.503.2380S

However, its notable that the original paper has 1578 citations.

23 New Methods

I didn't have to learn any new Python methods, but I used `%pip show seaborn` to show information like version number and dependencies, which I don't believe was covered in class but I already knew how to do it.

24 Final Disclaimer

I've used some aspects of seaborn before for research papers or independent computer science optimization problems I've worked on, but those were restricted to plot forms that I probably could've also done with matplotlib (e.g. scatterplots, regressions), so this is the first time I've looked at things like violin plots (in my Jupyter Notebook) and pair plots (in the papers from the ASCL Record).